PCT/GB 2004 / U U 1 5 4 5

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

04 JUN 2004
WIPO PCT

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 19 May 2004

EFFECTIVE FORM

1/77

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
CF
17 APR 2003
RECEIVED BY FAX

The Patent Office

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
South Wales

17APR03 E801202-1 C01F18 800
P01/7700 0.00-0308943.0

1. Your reference — NEW/RJ

2. Patent application number *(The Patent Office will fill in this part)* — 0308943.0 — 17 APR 2003

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*
The University of Dundee
Nethergate
Dundee DD1 4HN

Patents ADP number *(if you know it)* — 891127003

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention — A system for determining the body pose of a person from images

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

~~Ron Jenkins~~
~~Research & Innovation Services~~
~~University of Dundee~~
~~11 Perth Road~~
~~Dundee DD1 4HN~~

KENNEDYS PATENT AGENCY LTD
FLOOR 5, QUEEN'S HOUSE
29 ST VINCENT PLACE
GLASGOW G1 2DT

51/77
4/3/04.

Patents ADP number *(if you know it)* — 08038 4002 ~~7982007003~~

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:* — Yes
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	11 /
Claim(s)	CF
Abstract	
Drawing(s)	

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Date 17/04/03

12. Name and daytime telephone number of person to contact in the United Kingdom — Rm Fanham 01382 348141

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

A system for determining the body pose of a person from images

Timothy J Roberts, Stephen J McKenna and Ian W Ricketts
Division of Applied Computing, University of Dundee
9 April 2003

Field of Invention

A system is described that takes as input a live or pre-recorded image in which a person is present and computes estimates of the position and orientation of each of the person's major body parts that are visible in that image.

Background

The system can be used to locate and estimate the pose of a person in a single monocular image. Alternatively, it can be used during tracking of the person in a sequence of images by combining it with a temporal prior propagated from other images in the sequence. In this case, it allows tracking to reinitialise after partial or full occlusion or after tracking of certain body parts fails temporarily for some other reason. Alternatively, it can be used in a multi-camera system to estimate the person's pose from several views captured simultaneously.

Many applications follow from this ability to determine body pose. The pose parameters determined can be used as control inputs to drive a computer game or some other motion-driven or gesture-driven human-computer interface. Alternatively, the pose parameters can be used to control computer graphics, perhaps an avatar. The pose of a person can be used in the context of an art installation or a museum installation to enable the installation to respond interactively to the person's body movements. The detection and pose estimation of people in video images in particular can be used as part of automated monitoring and surveillance applications such as security or care of the elderly. The system could be used as part of a markerless motion-capture system with all the applications that has such as animation for entertainment and gait analysis. In particular, it could be used to analyse golf swings. The system could be used to analyse image/video archives or as part of an image indexing system.

Summary of Invention

The system has the following essential features:

(i) *Body part detectors*, one for each body part such as an upper arm, a lower leg or a head. These determine how likely it is that an individual body part is present at a given position, orientation and scale in the image with a particular shape and elongation.

(ii) A mechanism for scoring a hypothesised configuration of body parts *as a whole*. A configuration may consist of a single part (e.g. a head), two parts (e.g. an upper and lower arm), three parts (e.g. a hand, an upper arm and a lower leg) or indeed any number of parts up to and including the entire body. The configurations are scored in such a way that the scores of

configurations with different numbers of parts can be compared fairly. The score takes into account self-occlusion. This and the ability to compare configurations of different dimensionality are important features of the invention. The score takes into account the outputs of the body part detectors as well as other prior knowledge about symmetry (e.g. the lower right arm is likely to have a similar colour to the lower left arm) and visibility.

(iii) A population-based search routine which uses the outputs of the body part detectors to hypothesise configurations, score these configurations and search iteratively for probable configurations. Thereby, the search routine is able to 'bootstrap' from the outputs of the individual body part detectors to find configurations with more and more parts, ultimately finding a configuration comprised of all the parts that are clearly visible in the image. Information about the expected body pose can then be given as output from the system to control a particular application domain.

The accompanying paper which will be submitted for publication describes in more detail the background to the invention. It compares it with existing work and it describes an embodiment of the invention.

It will be clear to those skilled in the art that modifications and alternatives can be practiced within the spirit of the invention.

For example,

- the use of histograms could be replaced by some other method of estimating a frequency distribution (e.g. mixture models, Parzen windows) or feature representation.
- different methods for comparing feature representations could be used (e.g. chi-squared, histogram intersection)
- the part detectors could use other features (e.g. responses of local filters such as gradient filters, Gaussian derivatives or Gabor functions)
- the parts could be parameterised to model perspective projection (as opposed to the affine model adopted here)
- the search over configurations could incorporate any number of the widely known methods for high-dimensional search instead of or in combination with the methods mentioned above
- the population-based search could use any number of heuristics to help bootstrap the search (e.g. background subtraction, skin colour or other prior appearance models, change/motion detection).

Human Pose Estimation from Real-World Images

Timothy J. Roberts, Stephen J. McKenna, Ian W. Ricketts
Department of Applied Computing
University of Dundee
Dundee, Scotland, DD1 4HN
troberts@computing.dundee.ac.uk
www.computing.dundee.ac.uk/staff/troberts

Abstract

This paper presents a system for human pose estimation from real-world images. Its success is due to the synergy of three novel components. Firstly, a formulation is presented that models other object occlusion and thereby allows the representation and comparison of partial (lower dimensional) solutions. Secondly, a part detection scheme is described which, by using the high-level shape model earlier in the detection process, is able to account for texture. This results in more accurate part detection in real-world images. Furthermore, the response is less sparse than the popular edge approach, allowing more coarse sampling. Finally, a heuristic search scheme is discussed which takes advantage of these properties by firstly identifying possible body parts and then iteratively combining these partial solutions and predicting larger assemblies to achieve pose estimation without making assumptions about self occlusion.

1 Introduction

The aim of the system presented here is to recover a detailed estimate of human pose from real-world images, such as those produced by human computer interfaces or video indexing applications. This problem is made difficult primarily by the large variation in human appearance. Sources of variation include scale, viewpoint, surface texture, illumination, self-occlusion, other object occlusion, body structure and clothing shape. To account for this large and complex appearance it is popular to use a high-level, hand-built shape model (see for example the survey papers [5, 8]). Then, by associating points on a shape model with image measurements to compute a score, a search is performed to find the best solution(s). The success of this approach, in terms of its accuracy and efficiency, depends critically on the choice of model and its implicit assumptions. To achieve detailed, efficient human pose estimation from real-world images requires a system with components that work together synergistically. For example, the likelihood formulation cannot ignore the resulting estimation problem.

1.1 Outline

This paper describes three key, novel ideas. Firstly, a formulation is described that allows the representation of general occlusion and the comparison of partial solutions. Secondly, a part detection scheme is presented that incorporates the high-level shape model earlier in the detection process thus making part detection of in real-world images easier and more accurate. Thirdly, a two-stage heuristic search strategy is discussed that takes advantage of the properties of the induced space by identifying parts and then iteratively combining into larger configurations without making assumptions about self-occlusion.

Background material is discussed throughout the paper where most relevant. Due to the obvious similarities between pose estimation and tracking and since more work has been performed on tracking, references are made to both bodies of literature.

2 Body Model

A part-based representation is used to model the body shape. Each part has a set of pose parameters, which are to be recovered. A particular pose is then described by the parameters for each part, $P = \{p_l\}_{l=1}^{N}$, where l labels the body part and $N = 14$.

2.1 Part Visibility

A key assumption of current pose estimation and tracking systems is that all parts are visible. However, since parts may be occluded by other objects or difficult to localise, for example due to poor illumination, the system should be capable of representing a pose with some parts occluded. In this system, occluded parts are not parameterised. A consequence is that parts must be parameterised in their own co-ordinate system, rather than hierarchically as is often the case in tracking system, e.g. [15]. Whilst this increases the dimensionality of the space, in practice an offset term is often required to model complex joints like the shoulder [14] making the difference one of convenience. A pose is then described by the set of visible parts only.

For this parameterisation to be useful, a method such as that described in § 2.3 is required to compare poses with different numbers of visible parts. A key advantage of this approach is that it also allows the representation of *partial* (i.e. lower dimensional) solutions. This is important for two reasons. Firstly, due to their appearance, some parts might be found more easily than others. For example, it is often easier to locate parts that do not overlap. Secondly, because of inter-part linking, configurations with small numbers of parts contain much of the overall pose information. For example, knowing the position of just the head, hands and feet greatly constrains the pose.

2.2 Part Shape

Current systems often use geometric primitives, such as cylinders, to represent body parts [3, 15]. In contrast, here a non-parametric model of shape is learnt from manually segmented and aligned training images. Although each part could be represented using a distribution of masks, in this particular implementation each part is represented using a single mask. A point in such a mask, $m_l(x,y)$, represents the probability of membership

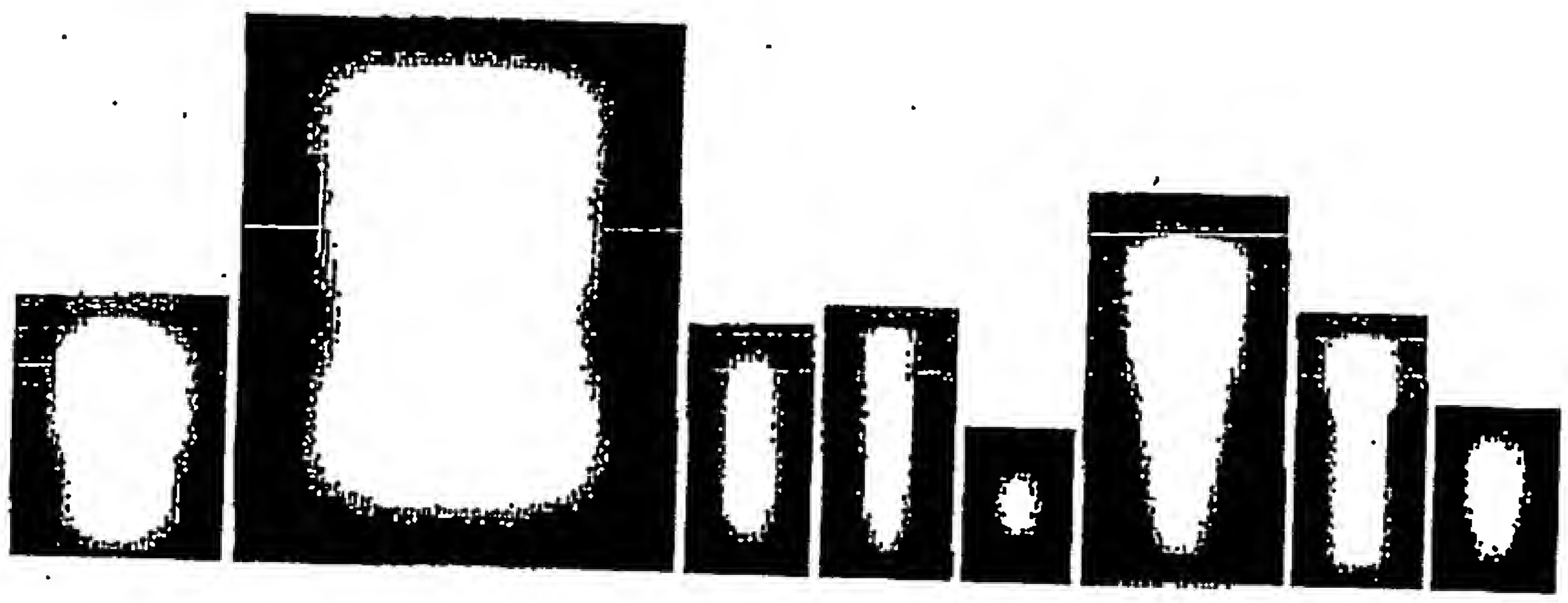

Figure 1: The learnt structure masks with major axis rotation marginalised.

to that part and is found by marginalising over the un-parameterised shape variations in the training data.

Due the limited presence of perspective effects, a variation of the scaled prismatic model [1] is used to parameterise the unaligned appearance. This reduces dimensionality compared to a 3D model and removes kinematic singularities [4]. In this case, rotation about the major axis is also marginalised since it is difficult to observe. These approximations are further justified due to the insensitivity of the part detector, as described below, to exact boundary position. The structure masks used throughout the paper were calculated from a small set of representative images from various subjects and are shown in Figure 1.

In this implementation, each part has a centre, (xc, yc), an image plane rotation θ and a foreshortening parameter e. Every part also shares a common scale parameter s. To account for self-occlusion, depth ordering is represented. The pose space is then denoted by the depth ordered set $P = \{p_i\}$ where $p_i = (xc_i, yc_i, \theta_i, e_i, s)$.

2.3 Part Appearance

Several methods for body part detection have been proposed although in the opinion of the authors much work remains to be done. Matching geometric primitives to an edge field is popular, e.g. [2, 15]. Sidenbladh *et al.* [13] emphasised the importance of learning the distribution of foreground and background local filter responses. Ronfard *et al.* [12] learnt part detectors from Gaussian derivative filters. However, in order to account for texture, less localised responses need to be computed. The use of highly local filters results in false positives (finding boundaries when none exist) and false negatives (failing to find a texture boundary). This is particularly important in the case of overlapping, textured clothing. Furthermore, matching local filters responses to a part boundary gives a sparse cue which necessitates dense sampling. Another popular method is modelling the background, but this has the obvious limitation of requiring knowledge of the empty scene. Finally, low-level texture region segmentation has immediate appeal but requires further investigation and no one-to-one correspondence exists between regions and parts.

In order to account for texture and give a less sparse response this system uses the high-level shape model earlier in the part detection process. Our approach is a general alternative to using local edge responses and has obvious similarities to model driven tex-

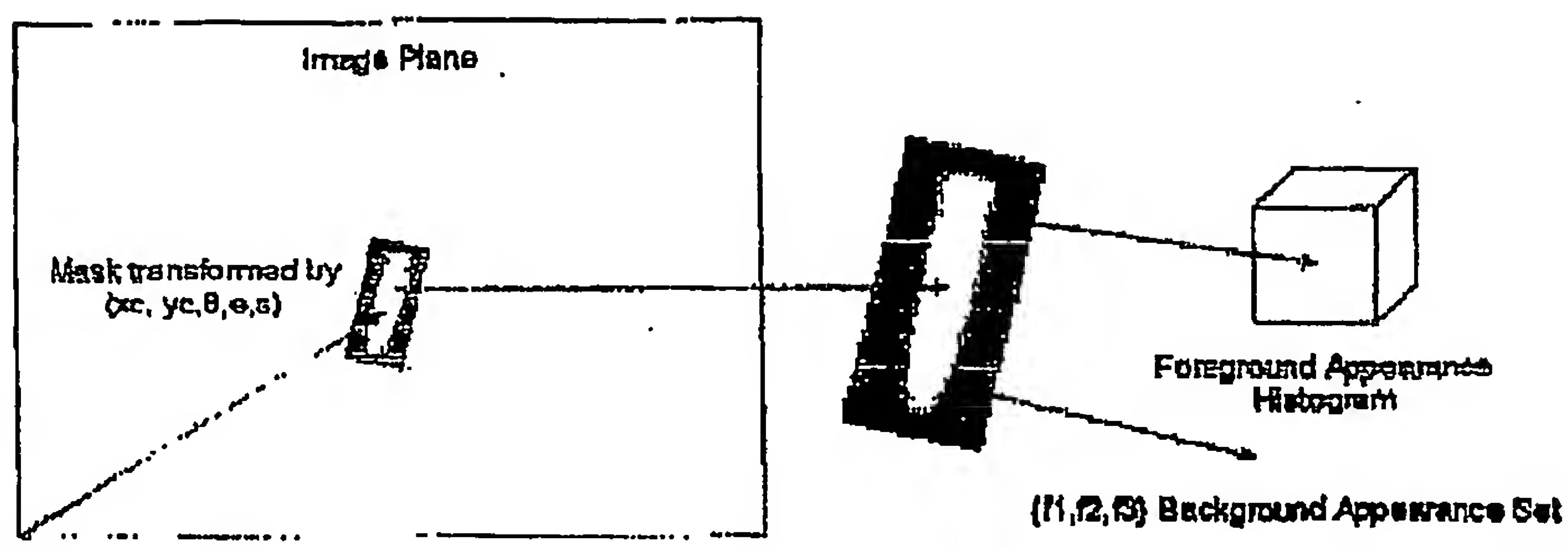

Figure 2: Illustration of the body model. The masks are transformed into image coordinates to collect the foreground and background appearance of the hypothesised pose.

ture segmentation. Arguments are now presented on what information is most important in formulating a part detector that is (i) feasible to learn and (ii) quick to evaluate. Firstly, it is assumed that neither the part foregrounds or the adjoining backgrounds are known. Next, it is argued that in general the spatial structure of the foreground and background is sufficiently unconstrained that it contains little discriminatory information. Whilst this is an obvious approximation, the computational cost of recovering this information is prohibitive. For example, describing the wide variation of layouts of clothing on a human torso would require a high dimensional model and large amounts of training data. The remaining information is present in the *differences* between the internal and external regions and the *similarity* of the internal regions.

To collect the foreground and background appearances for a hypothesised pose the structure mask for each visible part is projected into the image as shown in Figure 2. When taken together the masks determine the probability that a particular image feature belongs to a particular part or the background. A single point may contribute to multiple part appearances and the background appearance. More specifically, the probability that a feature $f(x,y)$ belongs to part i is $w_i(x,y) = m_i(x,y) \times \prod_j (1 - m_j(x,y))$ where j labels closer, visible parts. By extension, the background probability is $b(x,y) = \prod_k (1 - m_k(x,y))$ where k labels all visible parts.

The foreground appearance of each part is described using a single distribution. In this particular implementation, a feature, $f(x,y)$, is the intensity and chromaticity at the pixel, resulting in a limited model of texture. Standard histograms are used to represent the possibly multi-modal distributions. Histograms for each part, H_i are formed by adding features proportional to their weight $w_i(x,y)$.

It is proposed that the detection of a particular part should depend only upon regions close to that part. Therefore, background features are collected for each part and only from the region surrounding that part. A single feature can be shared between the background of multiple parts. In early implementations the background appearance was also represented using a feature distribution, but this required a large region of support that may not be present. Therefore, the background for a particular part is represented by a set of weighted features, $B_i = (b(x,y), f(x,y))$, where (x,y) are close to the mask.

2.4 Part Detection

A measure is required that allows the comparison of configurations with different numbers of visible parts. To accomplish this using a probabilistic formulation would require the evaluation of the model evidence, a high dimensional integration which is clearly not feasible. Ioffe and Forsyth [7] discuss some of the difficulties of probabilistic modelling of multiple part systems. Rather than insisting upon probabilistic modelling, un-normalised response strength is used instead, with response $R > 1$ signifying the presence of a configuration and larger response indicating greater support for the configuration. Our formulation naturally rewards higher dimensional configurations. The response is composed of a boundary response, a symmetry response and a link term:

$$R = \prod_{parts} RB_i \times \prod_{symmetric pairs} RS_{i,j} \times \prod_{neighbours} L_{i,j} \tag{1}$$

2.4.1 Boundary Response

The boundary response for a part is computed from the difference between its foreground and background appearances formed from the high-level shape model. More specifically, the response is a function of the probability of each border point not being foreground:

$$RB_i = F(\{D(x,y)\}) \tag{2}$$

$$D(x,y) = h(x,y) * (1 - H_i(f(x,y))) \tag{3}$$

The optimal form of this function is the ratio of the probability densities of a part given the response, to the probability of background given the response [13]. These probability densities are represented non-parametrically and learnt for each part from training images. In this implementation, to reduce the dimensionality of the distribution, background points are grouped together by proximity into segments. Segments are specified manually to account for the tendency for some segments to be stronger than others and reduce the bias for certain degrees of freedom in elongated shapes. The response for a segment is the average of the response of its points and segment responses are combined in an ad hoc fashion by assuming independence:

$$RB_i' = \frac{p(on|D')}{p(off|D')} \tag{4}$$

$$D' = \prod_{segments} \frac{\sum_{segment points} D(x,y)}{N_{segment points}} \tag{5}$$

2.4.2 Symmetry Response

If it can be assumed that the subject has body parts for which the foreground appearances are approximately equal, an additional term can be included to reduce the number of false maxima and increase accuracy. This is denoted by $RS_{i,j} = F(H_i, H_j)$. In the current implementation, histograms are compared using the Bhattacharyya measure,

$BH(H_i, H_j) = \sum_f \sqrt{H_i(f) \times H_j(f)}$. The Bhattacharyya measure for symmetric pair configurations and part background configurations is learnt in a similar way to the border response.

2.4.3 Link Term

To penalise invalid poses a link term is included. Each part has a set of control points that link it to its neighbouring parts. A link is introduced between each neighbouring part pair and takes the value $L_{i,j} = 1$ if the distance between the control points of the pair, δ, is less than the un-penalised distance, $\Delta_{i,j}$, and $L_{i,j} = e^{(\delta - \Delta_{i,j})/\sigma}$ otherwise. If the neighbouring parts do not link directly, because intervening parts are not parameterised, the un-penalised distance is found by summing the un-penalised distances over the complete chain. This can be interpreted as a force between parts equivalent to a telescopic rod with a spring on each end.

3 Pose Estimation

One can identify two distinct pose estimation strategies in the literature, one used primarily for single images and the other for tracking. Single image pose estimation schemes often use a two-stage approach e.g. [7, 12]. First, candidate body parts are detected in the image. Grouping is then performed, for example by dynamic programming, to identify compatible parts and the best possible solution. Whilst this strategy is efficient and global, the feed-forward nature does not address the problem of partial occlusion. Frame by frame trackers on the other hand proceed by localised sampling in the full dimensional pose space, e.g. [6], around estimates given by a motion model, e.g. [10]. The approach usually requires manual initialisation and does not recover from significant errors in the motion model. However, by sampling in the full dimensional space it does model self occlusion. Tracking systems often propose sophisticated search heuristics, e.g. [2, 14]. Of particular note is the application of the genetic search technique to a *hierarchical* pose space described in [3]. In the opinion of the authors this is make little sense as one cannot extract meaningful sub-solutions from the state (the motivation for genetic search) since each part is defined relative to its parents. The success of the approach could be due to a relatively small spread in the samples.

Due to problems in formulating the system using probabilistic logic, time constraints and the quantities of interest (i.e. maxima) this paper emphasises finding maxima over recovering an unbiased density estimate [9]. The key to efficient global optimization is in the use of heuristics that capture the important properties of the space. With this in mind, a two phase approach is proposed to take advantage of the formulation described above.

3.1 Population Phase

First, a global, low dimensional (often single part), 'population' phase is used to seed the search. This could be accomplished using a range of domain specific heuristics including motion, estimated foreground appearance (e.g. skin colour), estimated background appearance, context, or combinations thereof. In this particular implementation, due to the broad nature of the response, it is possible to perform coarse regular sampling over an expected region. The spacing of the sampling is determined manually on a part by part

basis. For example, the head is sampled horizontally and vertically every 5 pixels (for an image resolution of 640×480) and 8 orientations. Samples with a response larger than a threshold are added to the population.

3.2 Building Phase

Next, a 'building' phase is repeated for a fixed length of time to find larger configurations. For each iteration a new generation, of fixed size, is created and evaluated. The best assembly is preserved by adding it to each new generation (the elitest approach). In this particular implementation the building phase consists of a prediction step, a cross over step and a local search step.

In the prediction step, new samples are created by adding parts that are currently not parameterised into the set. This is necessary when the population phase failed to find the part, for example due to partial occlusion. Currently, this is performed in an ad-hoc fashion by adding a single part to the assembly for which the closest neighbour is currently parameterised. Parts are instantiated using the minimum control point distance, δ, to the closest neighbour, with a extension sampled from a Gaussian distribution (mean equal to 0.75, standard deviation=0.15), aligned with the closest neighbour and at the lowest depth. Currently, this step produces 25% of the new population.

In the cross over step, new samples are formed from the previous samples by combining body parts. The novelty is the improvement in which parts are sampled. Firstly, an assembly is randomly sampled from the previous population proportional to its response, R. Then a single part is sampled from that assembly proportional to its border response, RB_i. This term represents its fitness independent of the complete assembly. A complete assembly is produced by iterating this process, drawing a parts from the set of remaining parts in sampled assemblies, producing an assembly of size, N_G. Currently, N_G is chosen in an ad-hoc fashion as a sample from a Gaussian with a mean equal to the size of the current best assembly (elite) plus 1 and standard deviation of 1. This step produces the remaining 75% of the new population.

In the local search step, each member of the resulting population is improved by making local samples for each part in its assembly. More specifically, a part is sampled from the assembly and a small random perturbation, determined on a part by part basis, is made to the configuration. This perturbation includes a change in depth order. The new configuration replaces the old if it has a larger response. This step is iterated until no improvement is made for a fixed number of iterations (currently 20).

4 Summary

A system was presented that allowed detailed, efficient estimation of human pose from real-world images. The improvements were due to the synergy of three novel components: (i) a model of other object occlusion that also allowed the representation and comparison of partial (lower dimensional) solutions, (ii) a novel part detector that, by making use of high-level shape knowledge earlier in the detection process, is able to account for texture and thereby find body parts more easily and accurately (iii) a search scheme that takes advantage of these properties, firstly by identifying body parts and then by iteratively combining and predicting larger configurations, without making assumptions about self occlusion, to achieve efficient pose estimation.

References

[1] T. J. Cham and J. M. Rehg. A multiple hypothesis approach to figure tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 239–245, Fort Collins, Colorado, USA, 1999.

[2] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 126–133, South Carolina, USA, 2000.

[3] J. Deutscher, A. Davison, and I. Reid. Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 669–676, Hawaii, 2001.

[4] J. Deutscher, B. North, B. Bascle, and A. Blake. Tracking through singularities and discontinuities by random sampling. In *IEEE International Conference on Computer Vision*, pages 1144–1149, September 1999.

[5] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, January 1999.

[6] D. M. Gavrila and L. S. Davis. 3D model-based tracking of humans in action: A multi-view approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 73–80, San Francisco, USA, 1996.

[7] S. Ioffe and D. A. Forsyth. Finding people by sampling. In *IEEE International Conference on Computer Vision*, 1999.

[8] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, March 2001.

[9] R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, University of Toronto, 1993.

[10] V. Pavlovic, J. M. Rehg, T. J. Cham, and K. Murphy. A dynamic Bayesian network approach to figure tracking using learned dynamic models. In *IEEE International Conference on Computer Vision*, pages 94–101, 1999.

[11] T. J. Roberts, S. J. McKenna, and I. W. Ricketts. Adaptive learning of statistical appearance models for 3D human tracking. In *British Machine Vision Conference*, pages 333–342, Cardiff, 2002.

[12] R. Ronfard, C. Schud, and Bill Triggs. Learning to parse pictures of people. In *European Conference on Computer Vision*, pages 700–714, Copenhagen, 2002.

[13] H. Sidenbladh and M. J. Black. Learning image statistics for Bayesian tracking. In *IEEE International Conference on Computer Vision*, volume 2, pages 709–716, Vancouver, 2001.

[14] C. Sminchisescu and B. Triggs. Covariance scaled sampling for monocular 3D body tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 447–454, Hawaii, 2001.

[15] S. Wachter and H. H. Nagel. Tracking persons in monocular image sequences. *Computer Vision and Image Understanding*, 74(3):174–192, June 1999.